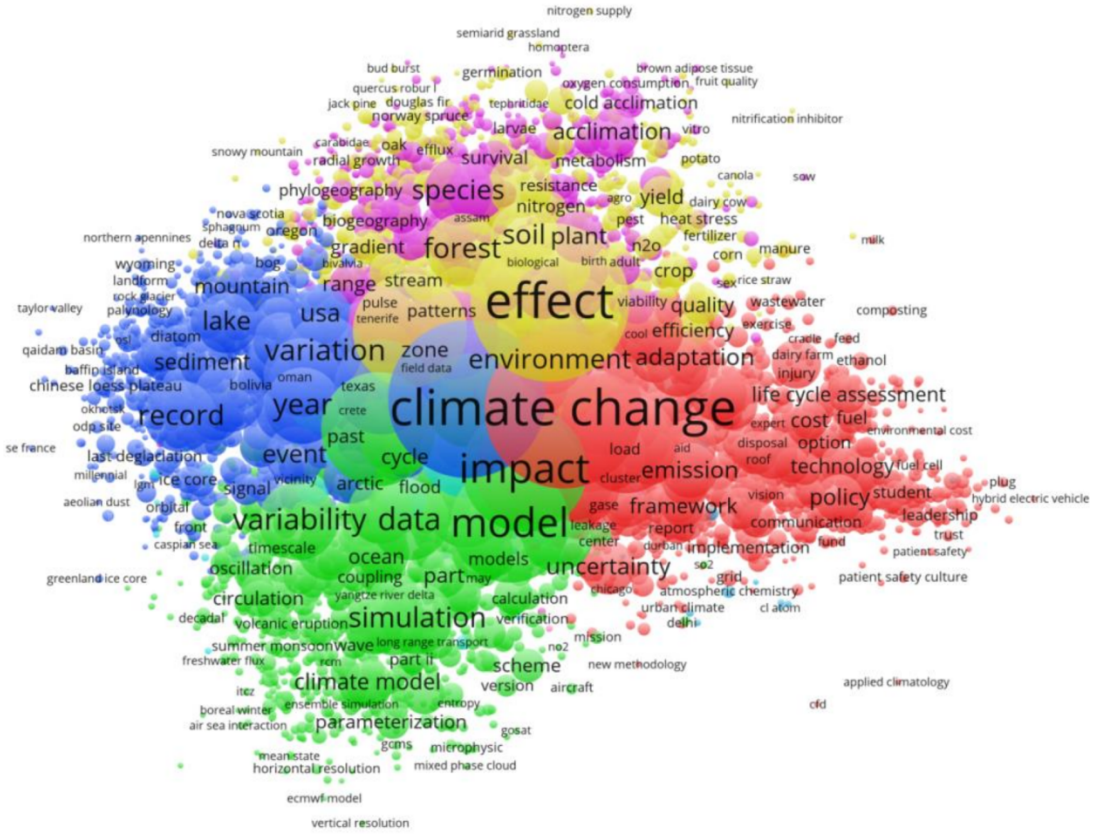
Accelerated Data Discovery for Scalable Climate Action

Henning Schwabe, Sumeet Sandhu,
Sergy Grebenshchikov

Presentation at the workshop "Tackling Climate Change with Machine Learning" at ICLR 2020.

Problem: Information Overload & Discoverability



Climate Data is Growing

There is exponential growth of information on climate change in the last 40 years - the number of peer reviewed papers doubles every 5 years.

Climate Data is Fragmented

Climate change is complex - its study spans natural sciences (meteorology, geoscience, chemistry, physics), ecosystem and economic modeling - with no standard "climate ontology."

New and Diverse Users

Unlike scientists modeling ecosystems decades into the future, new users in policy, health, insurance, information technology etc. seek either historical climate data or predictions 1-2 years out.

Climate Data to Climate Knowledge

Decision makers working on Climate Change Mitigation and Adaptation at local, regional, global levels need insights and knowledge - which must be distilled from disparate, highly technical, raw data sources.

1. Figure: R Haunschild, et al., "Climate Change Research in View of Bibliometrics," PLOS ONE, July 29, 2016
2. S. Tart, et al., "Market demand for climate services: An assessment of users' needs," Climate Services, Volume 17, January 2020
3. T. Kuramochi, et al., "Beyond national climate action: the impact of region, city, and business commitments on global greenhouse gas emissions," Climate Policy, 20:3, 275-291

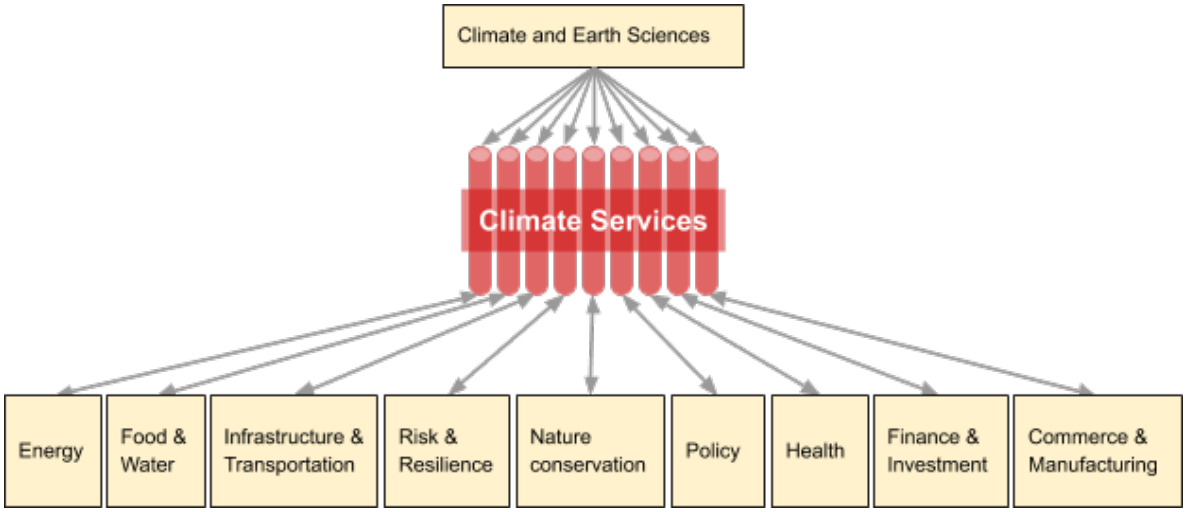
Existing Solutions: Climate Services

Climate Data to Knowledge - Today

Climate services offer the complex data transformation of natural and societal observations to practically actionable steps for Climate Change Mitigation and Community Adaptation. They consist of expert services and software tools - separately for each sub-domain.

Gaps in Climate Services

- Data is fragmented and siloed across disciplines and organizations
- Use cases aren't well defined - User Interfaces are ad-hoc, non-intuitive, too technical
- Databases are heterogeneous in coverage, quality, formats, structures, units, APIs, etc.
- There is no standard Climate Ontology across sources
- There is no standard metric for accuracy and veracity
- Local and hyper-local data and models for actionable timescales are missing - IPCC resolution is global over decades
- Climate Services are not easily scalable across the global diversity of economics, culture, impact, etc.



TYPE	REFERENCES
Climate Information Websites	• B. Hewitson, et al., "Climate information websites: an evolving landscape," Wiley Online Library 2017 • R. Swart, et al. "Developing climate information portals with users: Promises and pitfalls," Climate Services Journal 2017 • Clean Air Partnership "Scan of International Climate Information Portals", Environment and Climate Change Canada 2018
Climate Services	• https://www.sciencedirect.com/journal/climate-services • World Meteorological Organization, "Global Framework for Climate Services," 2009 • C. Vaughan, et al., "Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework," Wiley Interdisciplinary Reviews: Climate Change 2014 • European Commission, "A European research and innovation roadmap for climate services," 2015

Proposed Solution: Climate Catalog

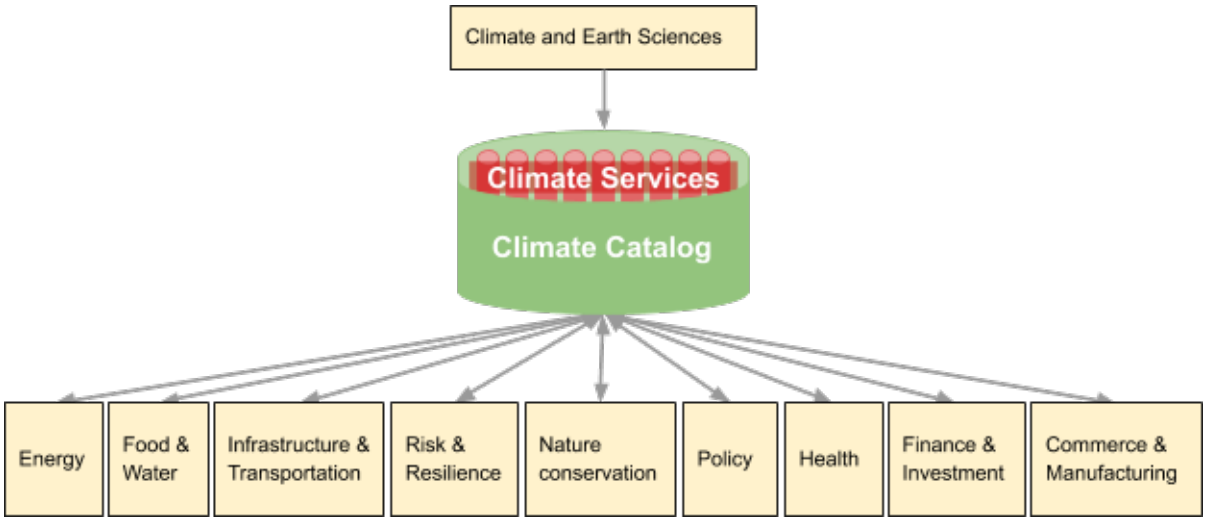
Manage Climate Data, Use Cases, and Solution Case Studies at Scale in Open Source

Climate Catalog

- Guide users to different data sources based on user roles and information needs
- Track new and emerging sources
- Identify whitespaces for new climate services
- Accelerate creation of new databases

Meta Data

- Development of a unified Climate Ontology
- Automatic summarization for quick discovery
- Explicit and implicit curation by climate services experts - data used in AI models for discovery
- Learn connections between data sources, use cases, and climate services - recommendations
- Open standard compatible with open source software



Use Case: Climate Policy Database

- Two databases are of interest to NGOs working on climate policy:
  - All national/regional climate-relevant policies under consideration; and
  - Companies' actions and positions on past/current climate-relevant policies.
- The challenges are:
  - Government information is often unwieldy and hard to navigate - doing this manually cannot meet the urgency of climate action.
  - Company climate-relevant information is not always public, and policy positions are often embedded in unstructured data.
- **Climate Catalog** can scale and accelerate the creation of these databases by training AI models with smaller, manually curated databases.

1. Example of data catalog: World Meterological Organization.
2. Example of ontology: L. McGibbney, "Semantic Web for Earth and Environmental Terminology (SWEET) Ontologies."

Contact

Authors

- Henning Schwabe henning_schwabe@icloud.com
- Sumeet Sandhu sumeet.k.sandhu@gmail.com
- Sergy Grebenshchikov: sgreben@gwdg.de

The authors would like to acknowledge helpful contributions from:

- Kameron Rodrigues kameronr@stanford.edu
- Rohan Nuttall rohan.nuttall1@gmail.com
- Cathy Chiba cmchiba@dauratus.com